

A COMPARATIVE ANALYSIS OF TRADITIONAL AND CLOUD DATA WAREHOUSE

KHAWAJA UBAID UR REHMAN¹, UMAIR AHMAD¹, SAJID MAHMOOD²

¹Department of Computer Science, University of Management and Technology, Lahore, Punjab, Pakistan

²Department of Informatics and Systems, University of Management and Technology, Lahore, Punjab, Pakistan

Email: ubaid.rehman@umt.edu.pk, 15026050017@umt.edu.pk, sajid.mahmood@umt.edu.pk

ABSTRACT. *In the age of emerging technologies, the amount of data is increasing very rapidly. With the passage of time, the methods of data handling are getting improved. Prediction analysis is quite a tough task, but it also yields interesting results. Different sectors like financial services, transportation, health and education are generating large amount of data. The emergence of web 2.0 (social web) made it possible for users and researchers to analyze and predict huge amount of data. The domain of Business Intelligence is core technology for users who want to extract useful information for decision making regarding their businesses. Data warehouse provides an insight into the business processes using the historical data. However, traditional data warehouse may not be suitable for the data analysis needs because of the evolving requirement of industry. It cannot be scaled up or down. Moreover, it cannot handle the increasing number of users. A new kind of data warehouse with design and implementation aspects has been emerged, called as cloud data warehouse. The cloud data warehouse model has evolved with the passage of time, which affects the application and business domains as well. The cloud data warehouse has evolved to control the large scale data. It can be scaled up or down at any time and also it has no limitation on increasing number of users. In this review paper, we have compared traditional and cloud data warehouse. We can conclude that the ultimate future of data warehouse is cloud data warehouse.*

Keywords: Data Warehouse (DWH); Traditional Data Warehouse (TDWH); Cloud Data Warehouse (CDWH); Online Analytical Processing (OLAP)

Introduction. Data warehousing is a standout amongst the most modern areas in the computing industry nowadays. For business directors, it has been given significant improvement to their business processes, while information system managers consider it high quality method to overwhelm the standard obstruction for presenting enterprise records for executives and other end customers. In computing, DWH is being used for reporting and data analysis, which is considered an essential factor of business intelligence. DWHs serve as the repository to store historical, sometimes current also, data from multiple sources in a single repository. The current data warehouse is more profitable and fascinating then its traditional data warehouse [1]. Bill Inmon quoted “A data warehouse is a subject-oriented, integrated, time varying, non-volatile collection of data that is used primarily in organizational decision making” [2]. The data warehouse supports OLAP, the efficacy and performance requirements unlike OLTP.

A traditional data warehouse provides full support of SQL, but traditional data warehouses are not scalable. It requires a lot of time to configure, optimize and manage the system. Nowadays organizations are moving towards cloud data warehouse. The idea of a cloud data warehouse was introduced for this. A cloud data warehouse is entirely different from traditional data warehouse. Using cloud data warehouse customers can get data from different time zones and geographic. The criteria for selecting a cloud data warehouse is

dynamic and elastic support of any scale of workload and user concurrency. A cloud environment is designed to take advantage of a larger number of user applications and it becomes possible for a DBA to keep up for that environment. So, the cloud data warehouse needs to take and provide new technology, techniques and automate a lot of these manual tasks.

Extract, Transform and Load (ETL) is the basic method used to populate a data warehouse and a core phase of data integration. The ETL is a back-end process which includes collecting data from multiple data sources; transforming it according to commercial needs, and stacking it into a data warehouse. Extraction is a procedure of reading data from multiple sources (homogenous/heterogeneous). The method of transforming the extracted data from its original state into stable state so that it can be used in later stages of the data warehouse is transformation. Most of the DWH building time, i.e. ETL process consumes approximately 70% of computing resources of data warehousing [3].

Business Intelligence (BI) is the core technology for users who want to deliver useful information for decision making. They pull out such information from the data warehouse. As we know, the transforming of the raw data into useful information is called Business Intelligence. We can use this information for better strategic and decision making tasks. Such methods can not only enhance applications in enterprise systems and industrial informatics, but it will also play a pivotal part to bridge enterprise systems and industrial informatics. In this review paper we have discussed the key differences between traditional and cloud data warehouse as shown in Table 1.

The paper structure is as follows: literature review is defined in the next section, critical analysis of traditional vs. cloud data warehouse is defined in section 3, discussion is defined in section 4, conclusion is defined in section 5 and references are given at last.

2. Literature Review: The latest technology of data warehousing provides solutions for businesses using streaming by mere means, connection of real time data with live data and historical data which is stored previously. New information using data warehousing takes into consideration which allowed information gathering and investigation crosswise over associations and offices.

There are many crucial prospects of data warehouse; cloud computing is one of them, due to its cloud-based computer architecture and multiple sources to access their data. The growing section in web investigation, research, development of distributed computing phases, administrations, which incorporate applications, programming framework, and equipment conveyed as administrations over the Internet, which is remarkable. Server virtualization, utility figuring, distributed computing has proposed programming as an administration [5].

The driving IT businessmen as of now pleased to bolster top of the line, which turn out BI & A (Business Intelligence and Analytics) applications utilizing distributed computing. For instance, amazon Elastic Compute Cloud (EC2) empowers clients to lease, virtual PCs on which to run their own particular PC applications. Its simple storage service (S3) gives online capacity a web benefit. Google app engine gives a stage to create and facilitate Java or Python-based web applications. The Big-table is utilized for backing information stockpiling in Google. The famous Windows Azure stage gives cloud administrations like SQL Azure and SharePoint permission of .Net system applications to keep running on the stage. The web and cloud administrations which drove business offer one of its kind information gathering, handling, and examination challenges for Business Intelligence & analytic individuals [5], [7].

The section of Business intelligence is an utterly different than the section of the conventional business room in any company, the data analytics took in different places and the Analysis is only limited to analyzing historical data.

Many businesses look inside into the historical data, so they can encounter the curve even before it's begun. Data Lakes provided by data warehousing new techniques and software which is different than storing data in historical files and folders happened in conventional data warehouses. On the other hand the flat architecture of data lakes allows un-processed data to be preserved in its natural form unless it's needed. The older Business Intelligence models line inefficient in terms of time and power consumed in data mobility and duplication. BI enhancement is the new kind of data warehousing. Most of the components are driving the growth and future of data warehousing. BI & A 1.0 advancements that are currently linked into a business try IT systems. Future BI & A 2.0 structures will need the coordination of creating flexible techniques in substance mining (e.g., information extraction, topic extraction, address answering), web mining, casual association examination, and spatial-transient examination with current DBMS-based BI & A 1.0 structures [6].

The analysts had concentrated on outlining adaptable frameworks for both fresh escalated workloads

and also unplanned examination workloads. Introductory plans incorporate circulated databases for refresh concentrated workloads, and parallel database frameworks for logical workloads. The parallel databases established past model frameworks to wide business frameworks, however dispersed database frameworks were not extremely fruitful and were never generalized instead of using many un-planned approaches to scaling. The altering in data access patterns of application and the required scale out to thousands of good machines headed to the birth of a new class of systems cited as key-value stores, which are now being broadly took over by many enterprises [8].

3. Comparative Analysis. Cloud data warehouse provides ‘storage as a service’. A traditional data warehouse requires lots of time to configure the hardware, software and infrastructure. It also takes a lot of time to optimize and manage the system. A cloud data warehouse is designed to take the advantage of a larger number of users and applications. Scaling up is a tedious and resource exhaustive task. The benefit of cloud DWH over traditional data warehouse is that scaling can be achieved effectively and easily. Using cloud, data can be scaled up or down instantly without any hassle on the other hand, in traditional data warehouse data cannot be scaled up and down instantly as the business needs changes. It is quick and easy to get a data warehouse up and running in cloud whereas deploying a traditional data warehouse takes a long time.

Cloud data warehouse are efficient for analytics because they use columnar storage and massively parallel processing (MPP) which allows better performance for running complex queries. Cloud data warehouse allows enterprises to move their focus from system management to analysis whereas traditional data warehouse focuses on data management. Cloud data warehouse provides cost benefits by eradicating hefty upfront costs. Using cloud data warehouse there are no hardware, server rooms, IT-related staffing issues, or operational expenses to maintain your DWH. Cloud data warehouse reduces the cost and complexity of managing on-premises systems so customers can focus on extracting value from their data rather than on maintaining hardware and software infrastructure.

So, cloud data warehouse provides new technology, techniques and automate a lot of these manual tasks whether that can be:

- Managing encryption.
- Automatic determination of data distribution.
- Delivering automatic performance optimizations.
- Deployment
- Limitation of users.
- Affordability
- Volume and pace of data
- Managing the data

The key differences between traditional data warehouse and cloud data warehouse are explained in Table 1.

Table 1. Traditional vs. Cloud DWH

Traditional Data warehouse	Cloud Data warehouse
Planning for traditional data warehouse can be a huge task.	You don't need to plan your data warehouse using cloud.
A traditional data warehouse is inflexible and it may result in over provisioning and overpaying.	A cloud data warehouse is flexible and automatically grows when its need increases. Using cloud data warehouse cost reduces.
A traditional data warehouse affect queries when the data grows.	A cloud data warehouse does not affect queries when the data grows.
A traditional data warehouse does not shrink when it is underutilized.	A cloud data warehouse automatically gets shrink when it is underutilized to save costs.
The criteria for selecting a traditional data warehouse is huge data volume but not increasing number of users.	The criteria for selecting a cloud data warehouse is dynamic and elastic data volume and increasing number of users
Using traditional data warehouse we	Data warehouse built for cloud, can be

cannot scale data up and down at any time. It takes hours or days to configure hardware, software and infrastructure. Also it cannot handle the increasing number of users.	“scale up and scale down at any time”. It will not take hours or days to configure hardware, software and infrastructure. It also handle increasing number of users.
A traditional data warehouse is expensive and not easy to scale when the data grows.	A Data warehouse built for the cloud is easy and inexpensive to size and scale.
A traditional data warehouse cannot handle diverse data types. It can only handle structured data.	Cloud data warehouse able to handle diverse data types i.e. semi structured and structured.
A traditional data warehouses forces you to buy compute and storage together.	Cloud data warehouse allows you to resize compute clusters (only pay for what you need, when you need it).

4. Dicsussion. With the cloud we are exploiting knowledge that our organization may not have. There are cost and operational benefits when working with the cloud supplier instead of doing it yourself. For instance, the typical procurement and purchase order process is tedious, making it problematic for a business to get computer systems when they need them.

The organizations can use cloud data warehouse to quickly scale up their systems. It provides automatic scaling of storage, analytics or workgroup resources for any job. As compared to traditional data warehouse cloud data warehouse services can easily be scaled up and down as the business needs change. Scaling is not an issue using cloud data warehouse. All the storage is handled by the cloud service provider for instance, Google Colossus file management system handles files, Google Borg cluster management system handle computing. It can also assigns jobs to hundreds of thousands of CPU's simultaneously.

Hadoop data lakes, other big data systems catch a considerable measure of attention and features nowadays, but the traditional data warehouse is still used by many organizations and for analysis of both recent and historical data. When we move towards cloud data warehousing the game changes. Using cloud data warehouse we don't need to worry about the cost and maintenance of data warehouses. With cloud data warehouses, maintenance becomes easy and the cost also reduces. Some cloud characteristics are shown in Figure-1.

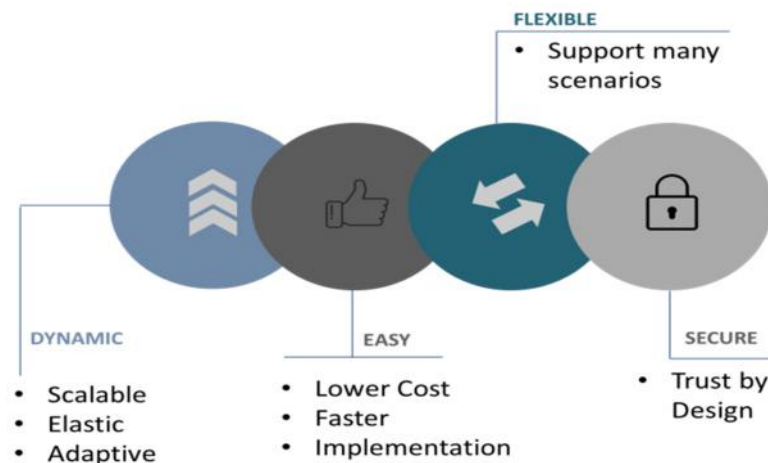


Figure-1: Cloud Characteristics

Company use predictive analytics to determine who should I contact to sell my products, what product enhancements, should I make, or simply check the statistics for example who will win the next football match?

A cloud warehouse allows you to process semi-structure data like JSON as well as structured data without transforming it first. JSON data is very popular for usage of logs from mobile apps, communication records, and post on social media. By processing JSON data in the warehouse directly using JavaScript and SQL, we can clean and prepare data much more quickly than with a model building framework.

A cloud warehouse allows us to clone when we want to build a predictive model. Cloning creates a snapshots and time data sets without duplicating storage, by using a static version of the data, we can optimize our model

and keep everything else constant. Cloning also allows you to create multiple copies of our data and test multiple predictive models at the same time. There is not interference between the models we create. Every model is built in separate versions of the data and cloud data should allow it to be like Time Travel app. The model of a cloud data warehouse is shown Figure-2.

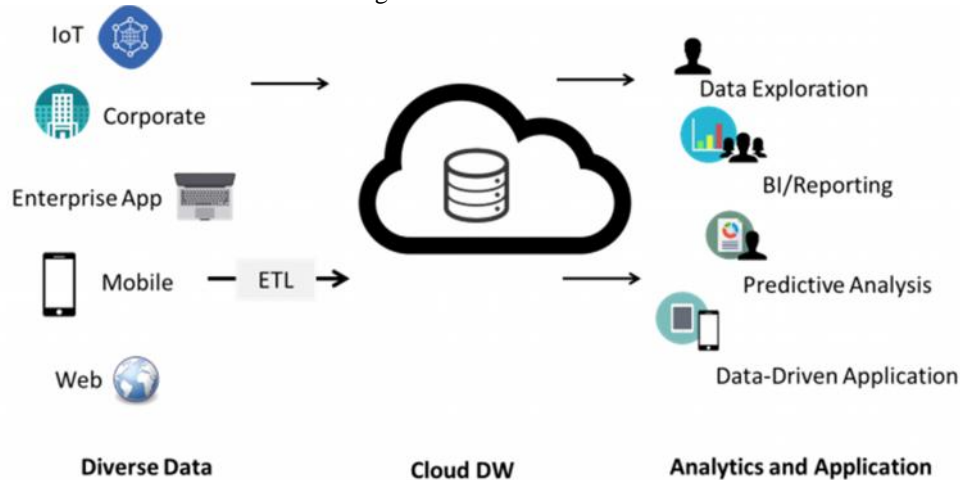


Figure-2: Cloud Data Warehouse

With Time Travel app, we can obtain a previous version of the data and compare it to the new versions of data. If our prediction turns out to be wrong, we can go back in time and analyze why it is so? We can also optimize the model parameters for better results next time. The Time Travel app allows us to recover any data accidentally deleted or updated. Essentially a cloud data warehouse offer us faster performance, more concurrency and more implicitly when doing data analytics.

For instance, Amazon web services data warehouse platform has built up the cloud as a feasible choice for organizations that need a quick and conceivably more cost-effective way to deploy data warehouses. The significance of cloud for data warehousing to Oracle clients is specifically connected to three key factors:

- A) Increased agility
- B) Better cost control
- C) Co-location

A. Improving agility

Using Oracle's cloud services data can be moved quickly and easily.

B. Better cost control

Using Oracle's enterprise cloud services it is easier to manage and more cost efficient in the light of the fact that the security features can be empowered by default and clearly improved and better.

C. Co-location for faster loading

Co-location offers speedier data loading which implies that business clients get more timely access to their data. Therefore, Oracle E-Business clients have now started moving their applications to Oracle's Cloud Services. [9], [10].

Conclusion. The modern data warehouse is changing the face of big data and business intelligence by providing an easy yet effective and powerful ways to achieve to the new trends. It's obvious to feel amazed by current speed and storage capacity at which the data can be retrieved and processed. The hardware and the performance are getting better with the passage of time.

The cloud data warehouse maintenance, scaling or upgrading is easier as compared to traditional data warehouse because in cloud data warehouse you do not need to maintain indexes, clean files or update Meta data etc. The cloud provides cheap storage and on-demand compute. The cloud data warehouse has lowered the cost of complexity and lengthy time-to-value, which has limited the adoption and successful use of traditional data warehouse technology. Moreover, cloud data warehouse provides fully independent scaling for compute, storage and services. In short, the cloud data warehouse is replacing the traditional warehouse as

predicting the main source of decision support and business analytics.

REFERENCES

- [1] Chaudhuri, S., & Dayal, U. (1997). An overview of data warehousing and OLAP technology. *ACM Sigmod record*, 26(1), 65-74.
- [2] Inmon, W. H. (2005). *Building the data warehouse*. John Wiley & sons.
- [3] El-Sappagh, S. H. A., Hendawi, A. M. A., & El Bastawissy, A. H. (2011). A proposed model for data warehouse ETL processes. *Journal of King Saud University-Computer and Information Sciences*, 23(2), 91-104.
- [4] Chaudhuri, S., Dayal, U., & Narasayya, V. (2011). An overview of business intelligence technology. *Communications of the ACM*, 54(8), 88-98.
- [5] Agrawal, D., Das, S., & El Abbadi, A. (2011, March). Big data and cloud computing: current state and future opportunities. In *Proceedings of the 14th International Conference on Extending Database Technology* (pp. 530-533). ACM.
- [6] Duan, L., & Da Xu, L. (2012). Business intelligence for enterprise systems: a survey. *IEEE Transactions on Industrial Informatics*, 8(3), 679-687.
- [7] Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: from big data to big impact. *MIS quarterly*, 1165-1188.
- [8] Cuzzocrea, A., Bellatreche, L., & Song, I. Y. (2013, October). Data warehousing and OLAP over big data: current challenges and future research directions. In *Proceedings of the sixteenth international workshop on Data warehousing and OLAP* (pp. 67-70). ACM.
- [9] Laker, K. (n.d.). (2017, March 03) Data Warehousing in the Cloud - Part 1. Retrieved from <https://blogs.oracle.com/datawarehousing/data-warehousing-in-the-cloud-part-1>
- [10] STEIER, T. (2018, May 25). To Cloud or Not to Cloud: Where Does Your Data Warehouse Belong? Retrieved from <https://www.wired.com/insights/2013/05/to-cloud-or-not-to-cloud-where-does-your-data-warehouse-belongs>
- [11] Widom, J. (1995, December). Research problems in data warehousing. In *Proceedings of the fourth international conference on Information and knowledge management* (pp. 25-30). ACM.
- [12] Abadi, D. J. (2009). Data management in the cloud: Limitations and opportunities. *IEEE Data Eng. Bull.*, 32(1), 3-12.
- [13] Guerhazi, E., Ayed, M. B., & Ben-Abdallah, H. (2015, June). Adaptive security for Cloud data warehouse as a service. In *Computer and Information Science (ICIS), 2015 IEEE/ACIS 14th International Conference on* (pp. 647-650). IEEE.
- [14] Nimje, A. R. (2015). Data Analytics as a Service (DAaaS): An Arriving Technology in Cloud Computing. *International Journal of Emerging Trend in Engineering and Basic Sciences*, 2(1).
- [15] Kurunji, S., Ge, T., Liu, B., & Chen, C. X. (2012, December). Communication cost optimization for cloud Data Warehouse queries. In *Cloud Computing Technology and Science (CloudCom), 2012 IEEE 4th International Conference on* (pp. 512-519). IEEE.
- [16] Lv, H. L., Wang, F. V., Van, A. M., & Cheng, V. L. (2012). Design of cloud data warehouse and its application in smart grid.
- [17] Abdelaziz, E., & Mohamed, O. (2015, December). Optimisation of the queries execution plan in cloud data warehouses. In *Information and Communication Technologies (WICT), 2015 5th World Congress on* (pp. 219-133). IEEE.
- [18] Sakr, S., Liu, A., Batista, D. M., & Alomari, M. (2011). A survey of large scale data management approaches in cloud environments. *IEEE Communications Surveys & Tutorials*, 13(3), 311-336.
- [19] Hussain, M., Al-Haiqi, A., Zaidan, A. A., Zaidan, B. B., Kiah, M., Iqbal, S., ... & Abdalnabi, M. (2018). A security framework for mHealth apps on Android platform. *Computers & Security*, 75, 191-217.

- [20] Iqbal, S., Kiah, M. L. M., Dhaghghi, B., Hussain, M., Khan, S., Khan, M. K., & Choo, K. K. R. (2016). On cloud security attacks: A taxonomy and intrusion detection and prevention as a service. *Journal of Network and Computer Applications*, 74, 98-120.
- [21] Khan, Y. D., Ahmad, F., & Anwar, M. W. (2012). A neuro-cognitive approach for iris recognition using back propagation. *World Applied Sciences Journal*, 16(5), 678-685.
- [22] Khan, Y. D., Khan, S. A., Ahmad, F., & Islam, S. (2014). Iris recognition using image moments and k-means algorithm. *The Scientific World Journal*, 2014.
- [23] Khan, Y. D., Ahmed, F., & Khan, S. A. (2014). Situation recognition using image moments and recurrent neural networks. *Neural Computing and Applications*, 24(7-8), 1519-1529.
- [24] Khan, Y. D., Khan, N. S., Farooq, S., Abid, A., Khan, S. A., Ahmad, F., & Mahmood, M. K. (2014). An Efficient Algorithm for Recognition of Human Actions. *The Scientific World Journal*, 2014.
- [25] Khan, Y. D., Abid, A., Farooq, M. S., Abid, K., & Farooq, U. A QUALITATIVE ANALYSIS OF FEATURE EXTRACTION BASED ACTION RECOGNITION TECHNIQUES.
- [26] Abid, A., Hussain, N., Abid, K., Ahmad, F., Farooq, M. S., Farooq, U., ... & Sabir, N. (2016). A survey on search results diversification techniques. *Neural Computing and Applications*, 27(5), 1207-1229.